R E P O R T R E S U M E S

ED 012 935 CG 000 422

DO COUNSELORS KNOW WHEN TO USE THEIR HEADS INSTEAD OF THE FORMULA.

BY- WATLEY, DONIVAN J.
NATIONAL MERIT SCHOLARSHIP CORP., EVANSTON, ILL.

PUB DATE 67

EDRS PRICE MF-$0.25 HC-$0.68 17P.

DESCRIPTORS- *PREDICTIVE ABILITY (TESTING), ACHIEVEMENT RATING, *GRADES (SCHOLASTIC), *COLLEGE STUDENTS, *COUNSELORS, *RESEARCH, CLINICAL DIAGNOSIS, STATISTICAL ANALYSIS, STUDENT RECORDS, COOPERATIVE ENGLISH TEST, UNIVERSITY OF MINNESOTA, MINNESOTA SCHOLASTIC APTITUDE TEST

THE PREDICTIVE SKILLS OF CLINICAL JUDGES WERE TESTED TO DETERMINE (1) IF VALIDATION EXPERIENCE AFFECTS THE ACCURACY OF CLINICAL JUDGMENT AND (2) IF THE CLINICAL JUDGE KNOWS WHEN TO DEVIATE FROM STATISTICAL PREDICTIONS. EIGHTEEN COUNSELORS WHO HAD PARTICIPATED IN A PREVIOUS INVESTIGATION OF PREDICTIVE SKILLS TOOK PART IN THE EXPERIMENT. ALL PARTICIPANTS WERE PROVIDED WITH INFORMATION REGARDING THEIR PREDICTIVE SKILLS IN THE PRIOR INVESTIGATION AND OTHER SPECIFIC DATA ABOUT CASE VARIABLES AND PRINCIPLES OF PREDICTION. THE PARTICIPANTS THEN WERE ASKED TO PREDICT FRESHMAN GRADES AND OVERALL COLLEGE GRADES FOR 50 CASES. CASE FOLDERS CONTAINED INFORMATION REGARDING SCHOLASTIC APTITUDE AND PAST ACADEMIC ACHIEVEMENT AS WELL AS STATISTICAL DATA SUCH AS EXPECTANCY TABLES. RESULTS INDICATED THAT THE PREDICTION OF FRESHMAN AND OVERALL COLLEGE GRADES DID NOT IMPROVE FOLLOWING THE VALIDATION EXPERIENCE. THE JUDGES FAILED TO INCREASE THEIR PREDICTIVE ACCURACY WHEN UTILIZING THEIR CLINICAL "SKILLS" RATHER THAN A STATISTICAL METHOD. THIS DOCUMENT IS A NATIONAL MERIT SCHOLARSHIP CORPORATION RESEARCH REPORT, VOLUME 3, NUMBER 1, 1967. (SK)

NMSC

ED012935

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

1967: volume 3, number 1

Do Counselors Know When to Use Their Heads Instead of the Formula?

Donivan J. Watley

NATIONAL MERIT SCHOLARSHIP CORPORATION

CG 000 422

Abstract

Two questions were investigated: (1) Does a general kind of validation experience improve the accuracy of clinical judgments? (2) Do clinical judges know when to use their heads instead of the formula? These questions were studied using judges known to predict educational criteria at relatively high, moderate, and low levels of accuracy. The results revealed that the accuracy of predictions of freshman and overall college grades did not improve after the validation experience; in fact, some evidence showed a decrease in accuracy. Further, the judges were clearly unable to improve predictive accuracy by attempting to recognize when to deviate from the formula.

Do Counselors Know When to Use Their Heads Instead of the Formula?[1]

Donivan J. Watley

Many questions remain unanswered in determining the relative efficiency of clinical and statistical methods of prediction. Answers were sought in this study to two questions specifically concerned with the predictive skill of clinical judges. The first relates to the argument of Holt (1958) and Gough (1962) that competitive clinical versus statistical prediction studies have not provided clinical judges with the same initial validation experiences available to the statistical method. That is, the statistical method is first developed on the same kind of sample and against the same criterion that is used in the comparative studies of the two predictive methods. Yet, the clinical judge typically is required to make predictions without having had any planned validation experience with the criterion prior to the competitive run. The present study provided clinical judges with one kind of prediction experience to determine whether this had any noticeable effect upon the accuracy of their forecasts.

The second question concerns Meehl's (1957) inquiry: When shall we use our heads instead of the formula? His analysis of a sizeable number of comparative clinical and statistical prediction studies led him to conclude that forecasts of outcome or institutional type criteria (e.g., college grades) will be more accurate in the long run when they are based on the actuarial method. Only in unusual circumstances should the clinical judge use his

---

[1] The data used in this study were collected while the author was on the staff of the Student Counseling Bureau, University of Minnesota, Minneapolis.

head (rely on his clinical "skills") rather than use the formula. Meehl suggests that the clinical judge use his head only when ". . . the psychological situation is as clear as a broken leg; otherwise, very very seldom" (1957, p. 273). But the important question remains: Does the clinical judge know when to deviate from the formula, i.e., recognize the "broken leg?"

Whether the clinical judge knows when to deviate from the formula is a question of considerable practical importance that surprisingly has received virtually no research attention. Since in the actual prediction situation the judge usually has the statistically derived prediction, if one is available, in addition to other case data, what really matters is whether the judge is able to use all of this information efficiently. The typical clinical versus statistical prediction study is designed unrealistically because the actuarial prediction itself is withheld from the clinical judge.

Method

Clinical Judges and the Validation Experience

Eighteen counselors took part in this study, all of whom participated in a previous investigation (Watley, 1966b) that assessed the predictive skill of individual counselors. A total of 66 high school and college counselors were in the first study and the 18 included in this study were specifically selected on the basis of their ability to predict: (1) freshman grades, (2) overall college grades, and (3) whether students would persist and be successful in the educational programs they selected at the time of admission to college.

Based on prediction records, the counselors were ranked from 1 to 66 on each of the three criteria. The two ranks for freshman and overall

grades were then combined, leaving one set of ranks for accuracy in forecasting grades and the other for judging persistence and graduation from initial educational programs. Counselors were identified who ranked in the top one-third (including ranks 1 to 22), in the middle one-third (ranks 23-44), and in the bottom one-third (ranks 45-66) on <u>each</u> of the two sets of rankings. Of the counselors identified at each level, six were randomly selected to participate in this study; and they were labeled respectively the high, moderate, and low accuracy groups. Use of these three groups made it possible to examine whether the validation experience was differentially related to the ability to predict accurately.

Prediction experience was acquired, therefore, in the first study. Predictions were made for the same sample of 100 cases in each of three conditions that differed in the type and amount of case information available. However, the judges were unaware that the same cases were included in each condition. The exact data provided in each condition can be found by referring to the initial study (Watley, 1966b).

The present study was conducted approximately one year after the first investigation. The following procedure was used to provide judges with further information about the prediction task. Approximately two months prior to this study each judge was given a report of the results obtained in the initial investigation (Watley & Vance, 1964). This report included information (listed by counselor identification number) about the number of correct predictions each judge made for each condition and the correlation coefficient between each judge's predictions and the grades actually obtained by students. In addition, specific data were provided about the case variables most highly related to the predicted criteria, as well as the differences in data typically used by judges who predict at relatively high,

moderate, and low levels of accuracy. Other information included: the relationship between counselor confidence in their judgments and actual predictive accuracy; the effect of place of employment (high school or college) on counselor predictive accuracy; the reliability of counselor judgments; and psychometric and biographic differences between counselors who predict educational criteria most or least accurately. About two days before making judgments in this study the judges were contacted and asked to review this material. The investigator then talked individually with each judge and two things were discussed: (1) the judge's performance in the first study and (2) information contained in the report that might generally be useful to improve predictive accuracy of grades. However, this was designed as a self-learning process in which information was provided but the judge was left to integrate it for himself.

The clinical judges predicted both freshman and overall college grades in this study. The effect of the validation experience was determined by comparing the number of correct predictions made in the initial study with the number made in the present study. A hit was defined as a correct dichotomized prediction for a student to earn a grade average of "C or higher" or "less than C," based on grades actually earned.

Deviation from the Formula

The judges were asked first to make freshman and overall college grade predictions for 50 cases. As indicated, this set of predictions was compared with predictions made in the earlier study (Watley, 1966b) to assess the effect of the validation experience. This set of predictions was also used to determine whether judges recognized when to deviate from the formula. After forecasts were made for all cases, the judges were then asked to go

back through each case folder again; only this time the statistical predictions for freshman and overall college grades were also available. The judge's job was to decide whether he should deviate from the statistical prediction in order to improve predictive accuracy. He was also aware of his first predictions for each case when the statistical predictions were not available.

Whether the judge recognized when to deviate from the formula was assessed in two ways: (1) the accuracy of his forecasts with and without the availability of the statistical predictions, and (2) the accuracy of his forecasts in comparison with the accuracy of statistical predictions.

The statistical predictions were cross-validated and were based on an equation that included high school rank (HSR), the Minnesota Scholastic Aptitude Test (MSAT) and the Cooperative English Test (CET).

Prediction Sample and Case Data

The sample was composed of 50 males who entered the College of Science, Literature, and the Arts (SLA) at the University of Minnesota as first-quarter freshmen in the fall of 1959. These students were randomly selected from among the entire entering class of freshman males. However, inclusion depended on the availability of all of the desired psychometric and biographic case data, graduation from a Minnesota high school during the spring of 1959, and at least one quarter spent in SLA.

Each case folder contained information related to scholastic aptitude and past academic achievement. Test scores were provided for the MSAT, the CET, and the Social Studies Test of the Sequential Tests of Educational Progress. Achievement data included each student's HSR and the last high school grades earned in the areas of mathematics, English, social studies,

and natural sciences. Also included were results for the Strong Vocational Interest Blank and the Minnesota Multiphasic Personality Inventory, plus considerable biographic information given on the Minnesota College Admissions Form and the Personal Inventory for Entering Students.

Statistical data were also provided to each judge for use in making predictions. This included: freshman grade expectancy tables for HSR, MSAT, and the CET; and a regression equation that included prediction coefficients for the high school grades of mathematics, English, social studies, and natural sciences.

The type and amount of case information provided in these folders corresponded to the third condition under which judgments were made in the initial study (Watley, 1966b). Essentially, these folders contained all of the data that were available for this group of students before they entered college. Therefore, the number of correct predictions in this study were compared with the number of hits made by judges in the third condition of the first investigation. However, since judgments were made for 100 cases in the first study and 50 in this one, the total number of correct forecasts obtained by each judge in the first study was divided by two in order to make the number of cases comparable for the two investigations.

Results and Discussion

Does Validation Experience Effect the Accuracy of Clinical Judgments?

Table 1 shows the mean number of correct forecasts made by the high, moderate, and low accuracy groups of judges both before and after the validation experience. An analysis of variance was computed separately for each predicted criterion.

The main concern of these analyses was whether significantly more hits

Table 1

Mean Number of Hits Obtained by Judges Before and After the Validation Experience

Validation Experience		Level of Predictive Skill					
		High		Moderate		Low	
		First year	O-A	First year	O-A	First year	O-A
Before	Mn	36.1	32.7	34.7	30.0	31.5	27.9
	SD	1.5	1.6	2.5	1.0	3.9	2.1
After	Mn	36.8	30.5	32.5	27.2	29.0	27.8
	SD	1.6	2.0	5.8	2.6	5.8	2.8

were obtained by the judges after the validation experience. The F found for assessing this difference for freshman grades was not significant at the .05 level. Table 1 shows that the most accurate judges obtained about the same mean number of hits after the validation experience, while the moderate and least accurate judges made slightly fewer hits. Thus, no evidence was obtained that the previous prediction experience and the feedback information the judges received aided in producing more accurate judgments.

As expected, however, the F of 13.17 obtained for assessing the differences among the means for the high, moderate, and low accuracy groups was significant beyond the .001 level. The interaction term was not significant at the .05 level.

For the overall college grade judgments, the obtained F of 5.19 for assessing the effect of the validation experience was significant at the .05 level. Surprisingly, however, opposite results occurred than might have been anticipated. Rather than improving accuracy, Table 1 shows that

the high and moderate level judges predicted less accurately after the validation experience.

The F of 18.70 for assessing the mean differences among judges who predict at high, moderate, and low levels of accuracy was significant beyond the .01 level. This was expected. The interaction term was not significant at the .05 level (F=2.48).

Obviously, the kind of validation experience provided judges in this study did not help improve their predictive ability. What this apparently means is that familiarity with general information that could be useful in improving predictive ability is not sufficient. Both Soskin (1954) and Crow (1957) found similar results to the extent that accuracy failed to improve under conditions that were not well defined. As was found here, Crow's judges were somewhat less accurate in interpersonal perception after training, a loss that seemed related to a decreased sensitivity to individual differences. In this study it is likely that some of the judges were unable to effectively integrate this new information, became somewhat confused, and predicted overall college grades less accurately than they would have without these data to synthesize.

Perhaps in addition to general information, a systemized form of immediate feedback after specific predictions would be more successful in building internal norms and, thus, help to improve the accuracy of clinical judgments of this type. Taft (1955) previously suggested this possibility and Oskamp's (1962) research demonstrated some success with this approach. However, the question then becomes: to what extent should one go in order to train clinical judges to predict institutional-type criteria as accurately as the equation can do already? Theoretically, specific training would

be necessary for every specific criterion. Perhaps the clinical judge's time would be better spent analyzing and improving his predictions of criteria for which the statistical method is not applicable.

Does the Judge Recognize When to Deviate from the Formula?

The first analysis was a comparison of the accuracy of judgments made with and without the availability of statistical predictions. The latter judgments were made with instructions to decide when to deviate from the formula, i.e., recognize the "broken leg" cases. The mean number of hits obtained by the judges under these two conditions are shown in Table 2.

Table 2

Mean Number of Hits Obtained by Judges in "Deviating from the Formula"

Availability of Statistical Predictions		Level of Predictive Skill					
		High		Moderate		Low	
		First year	O-A	First year	O-A	First year	O-A
Without	Mn	36.8	30.5	32.5	27.2	29.0	27.8
	SD	1.8	2.0	5.8	2.6	5.8	2.8
With	Mn	36.0	30.3	32.5	28.3	29.8	28.3
	SD	1.4	1.4	5.4	2.1	3.9	3.5

For freshman grades, the F for assessing the correct predictions made by judges under the two conditions was not significant. In fact, the total mean number of hits (32.8) for the three groups of judges was identical for both conditions. Thus, not only were the judges unable to effectively decide when to deviate from the formula, the statistical predictions had relatively little effect in any direction on the accuracy of their forecasts. The F of 13.57 for assessing the differences among the three accuracy groups

was significant beyond the .01 level; and the interaction term was not significant at the .05 level.

The results found for the overall college grade predictions were essentially the same. The F for assessing the differences with and without the statistical predictions available was not significant at the .05 level; but the F of 7.27 for the three accuracy groups was significant at the .05 level. Also noteworthy with this prediction is the fact that no differences were observed between the "moderate" and "low" level judges in the number of hits made. However, three criteria were used in the initial selection of the three accuracy groups and there was little variation among the judges in their ability to predict overall college grades.

The second analysis compared the number of hits made by judges when they attempted to recognize the "broken leg" cases with the number of correct predictions made by the actuarial method. The equation that included HSR, MSAT, and the CET correctly predicted "C or better" or "less than C" freshman grades for 35 cases and overall college grades for 31 cases. Table 2 shows that the most accurate judges were able to make forecasts of both criteria about as accurately as the statistical method. An analysis of their individual judgments showed that they tended to remain rather closely in agreement with the statistical predictions.

Judges who predicted at the moderate and lowest levels were inclined to deviate more frequently from the statistical predictions, preferring to remain in agreement with their initial judgments made without the statistical forecasts. As Table 2 shows, the availability of the statistical predictions had no noticeable effect on the accuracy of their judgments. Although demonstrating confidence in their predictions, this also reveals that the poorer judges failed to learn from the information provided to them earlier. For

example, they did not learn that judges who predict educational criteria least accurately tend to express more confidence in their forecasts than judges who predict most accurately (Watley, 1966a); or that they were more likely to improve predictive accuracy of institutional-type criteria by sticking rather closely to the statistically derived forecasts.

Thus the results obtained were disappointing. The judges who previously demonstrated the highest level of predictive ability were unable to improve on the accuracy of the statistical method by recognizing "broken leg" cases in which the statistical forecast was likely to be in error. However, the best judges tended to approach this task cautiously, unwilling to trust their judgment to select likely "deviate" cases. More alarming, however, is the fact that counselors in the moderate and low level groups stubbornly persisted in believing in the correctness of their own judgments in spite of rather powerful evidence to the contrary. In the final analysis, Meehl's warning is as appropriate as before except that in making forecasts of institutional criteria it seems that the judge should deviate from the formula "very, very, very seldom."

References

Crow, W. J. The effect of training upon accuracy and variability in interpersonal perception. Journal of Abnormal and Social Psychology, 1957, 55, 355-359.

Gough, H. G. Clinical versus statistical prediction in psychology. In L. J. Postman (Ed.), Psychology in the making. New York: Knopf, 1962.

Holt, R. R. Clinical and statistical prediction: A reformulation and some new data. Journal of Abnormal and Social Psychology, 1958, 56, 1-12.

Meehl, P. E. When shall we use our heads instead of the formula? Journal of Counseling Psychology, 1957, 4, 268-273.

Oskamp, S. The relationship of clinical experience and training methods to several criteria of clinical prediction. Psychological Monographs, 1962, 76 (28, Whole No. 547).

Soskin, W. F. Bias in postdiction from projective tests. Journal of Abnormal and Social Psychology, 1954, 49, 69-74.

Taft, R. The ability to judge people. Psychological Bulletin, 1955, 52, 1-23.

Watley, D. J. Counselor confidence in accuracy of predictions. Journal of Counseling Psychology, 1966, 13, 62-67. (a)

Watley, D. J. Counselor variability in making accurate predictions. Journal of Counseling Psychology, 1966, 13, 53-62. (b)

Watley, D. J., & Vance, F. L. Clinical versus actuarial prediction of college achievement and leadership activity. Final report, 1964, Project No. 2202, Cooperative Research Program, Office of Education, U. S. Department of Health, Education, and Welfare.

Previous NMSC Research Reports

Volume 1, 1965

Number

1. The Inheritance of General and Specific Ability, by R. C. Nichols.
2. Personality Change and the College, by R. C. Nichols (also in American Educational Research Journal, in press).
3. The Financial Status of Able Students, by R. C. Nichols (also in Science, 1965, 149, 1071-1074).
4. Progress of the Merit Scholar: An Eight-Year Followup, by R. C. Nichols and A. W. Astin (also in Personnel and Guidance Journal, 1966, 44, 673-686).
5. Prediction of College Performance of Superior Students, by R. J. Roberts.
6. Non-intellective Predictors of Achievement in College, by R. C. Nichols (also in Educational and Psychological Measurement, 1966, 26, 899-915).
7. Ninth Annual Review of Research, by the NMSC Research Staff (superseded by the Tenth Annual Review).
8. Social Class and Career Choice of College Freshmen, by C. E. Werts (also in Sociology of Education, 1966, 39, 74-85).

Volume 2, 1966

1. Participants in the 1965 NMSQT, by R. C. Nichols.
2. Participants in the National Achievement Scholarship Program for Negroes, by R. J. Roberts and R. C. Nichols.
3. Career Choice Patterns: Ability and Social Class, C. E. Werts.
4. Some Characteristics of Finalists in the 1966 National Achievement Scholarship Program, by W. S. Blumenfeld.
5. The Many Faces of Intelligence, by C. E. Werts (also in Journal of Educational Psychology, in press).
6. Sex Differences in College Attendance, by C. E. Werts.
7. Career Changes in College, by C. E. Werts (also in Sociology of Education, in press).
8. The Resemblance of Twins in Personality and Interests, by R. C. Nichols.
9. College Preferences of Eleventh Grade Students, by R. C. Nichols (also in College and University, in press).
10. The Origin and Development of Talent, by R. C. Nichols (also in Phi Delta Kappan, in press).
11. Tenth Annual Review of Research, by the NMSC Research Staff (includes abstracts of all previous NMSC studies).

NMSC Research is supported by grants from the National Science Foundation, the Carnegie Corporation of New York, and the Ford Foundation.